Analysis of Text-Semantics via Efficient Word Embedding using Variational Mode Decomposition

Rohith Ramakrishnan, Anirudh Vadakedath, U Vamsi Krishna, Premjith B and K.P Soman
Centre for Computational Engineering and Networking
Amrita School of Engineering Coimbatore
Amrita Vishwa Vidyapeetham
b_premjith@cb.amrita.edu

Abstract

In this paper, we propose a novel method which establishes a newborn relation between Signal Processing and Natural Language Processing (NLP) method via Variational Mode Decomposition (VMD). Unlike the modern Neural Network approaches for NLP which are complex and often masked from the end user, our approach involving Term Frequency - Inverse Document Frequency (TF-IDF) aided with VMD dials down the complexity retaining the performance with transparency. The performance in terms of Machine Learning based approaches and semantic relationships of words along with the methodology of the above mentioned approach is analyzed and discussed in this paper.

1 Introduction

Regardless of numerous years of research and development in Natural Language Processing, the models developed for various applications are primitive with limited linguistic knowledge. Over the years of dispute over how best to improve the NLP models, numerous approaches and techniques were proposed. During this period, the amount of text data for NLP research sky-rocketed. However, such an increase in the corpus didn't help NLP models developed for several languages and applications to escalate at the same rate.

Distinguishing the local and global scope of words is a crucial part of the NLP problems. It is because a word can possess different senses according to the contexts. The local scope of the word may be entirely different from its global scope. This increase in the ambiguity in perceiving the sense of words increases the complexity of the NLP tasks. In addition to these hassles, the major obstacle is deriving synonyms, semantic and lexical relationship between words. It has been a holy principle of NLP to evaluate the synonymity via annotation and semantical-labels. Even with the introduction of numerous neural network-based text classification such as BERT and fastText, the above-stated problems haven't been solved.

TF-IDF was one of the numerous techniques surfaced to generate a vectorial representation for the text data (Sparck Jones, 1972) (Barathi Ganesh et al., 2017). TF-IDF quantifies the significance of a word to a document (Ullman, 2011), by computing a weighting factor, which helps to identify the most frequently occurring words in the corpus. Another popular and most sought after model is fastText. It is a word embedding algorithm developed by Facebook's AI Reach lab (FAIR) for learning word embedding and text classification. fastText was able to retrieve character-level information for word representations and sentence classification. For example, considering the word "bunny" and n=3, the n-grams would be <bu, bun, unn, nny, ny>. The symbols < and > are added so that n-grams are not mixed with shorter words such as <bun> in this case. It helps to generate meaningful representations for suffixes and prefixes. Like fastText, GloVe (Global Vectors) is another model for distributed word representations. This model follows an unsupervised learning approach to get vector representations for words. GloVe is a hybrid approach that combines the concepts of count-based representation and distributed representation of words.

One of the uncharted sections of NLP is the association and application of Signal Processing concepts to improve and derive a solution to the above-mentioned problems. Over the past few decades, Signal Processing has gone to numerous heights giving us various tools such as Fourier, convolution and many recent methods such as Variational Mode Decomposition and Dynamic Mode Decomposition whose applications are numerous and unexplored.

In this paper, we introduce the application of one such concept from signal processing, Variational Mode Decomposition (Kalimuthu et al., 2019) to better solve the issue of synonymy and semantic relationship between words using basic Machine learning models. We discuss the existing methods and showcase the VMD based TF-IDF embeddings' performance with numerous basic Machine Learning models and its ability to better capture semantic relationships which are compared with existing neural network-based methods such as GloVe and fastText. Amidst the global pandemic of COVID-19,with enormous surge in the infection, people have continuously exhibited their sentiments on twitter. This corpus can be utilised to extract information relevant to understanding the consequences of Coronavirus.The proposed algorithm is also tested on the IMDb Data-set, consisting of people reviews labeled based on the sentiment. The results of the VMD based algorithm for the above-mentioned data-sets are discussed further in the Results and Discussion section.

2 Literature Review

With the arrival of word embedding models, a significant leaps have been achieved in the field of NLP. In (Ferrari et al., 2017) Alessio Ferrari, Beatrice Donati and Stefania Gnesi worked on detecting domain-specific ambiguities by seeing how the meaning of words in a specific domain like Computer Science would change when compared to the same word in other domains such as Mechanical Engineering, Medicine or Sports.(Rezaeinia et al., 2017)This paper proposes a novel method, Improved Word Vectors (IWV), which contributed to significance in the improvement of accuracy of sentiment based classification. Their experiment results show that Improved Word Vectors (IWV) are very effective for sentiment analysis . For a data-set of movie reviews, while GloVe and Word2Vec only shows 79% accuracy, IWV provide an accuracy of 80% and for the Stanford statement tree-bank data-set, while Word2Vec and GloVe methods showed 82% accuracy IWV showed 84% accuracy. (Balakrishnan et al.,) This paper uses Support Vectors for 2 tasks; task 1 is distinguishing the tweets mentioning "adverse drug reaction" from those do not and the task 2 is about distinguishing the tweets that include personal medication intake, possible medication intake and non-intake. (Soman and others, 2016) Dr.K.P Soman and others presented an approach using BoW (Bag of Words) and Neural Networks to classify questions consisting of English and Bengali texts. Application of theories from other sectors into NLP is not unusual. Zhou, Lei & Tse, K.T. & Li, Yutong have applied Higher Order Dynamic Mode Decomposition, which captures the eigen-frequencies, and fundamental transition dynamics and named their procedure EigenSent, that can represent well the sequence by generating an embeddings. (Zhou et al., 2021).

Variational Mode Decomposition being an fully intrinsic, novel and quasi-orthogonal decomposition method involving non-recursive extraction of modes (Dragomiretskiy and Zosso, 2013). VMD is the successor of the novel Empirical Mode Decomposition, which decomposes any given function into coarse multi resolution of Intrinsic Mode Functions. The simplicity and effectiveness of the algorithm had caught of lot of attention.EMD has extended its application in the field of engineering, mathematics and sciences. A plethora of documented literature is available on use of EMD in signal de-noising. In (Boudraa et al., 2006), Boudraa and Cexus have exhibited the application of EMD in de-noising signals. EMD can also be utilised for biomedical signal analysis (Weng et al., 2006). K. Khaldi and others exhibited the advantage of EMD in de-noising vocal signals (Khaldi et al., 2008), (Molla et al., 2013). Zhang et al introduced EMD based crude oil price analysis (Weng et al., 2006). Zhu further extended EMD with Genetic Algorithm for carbon price forecasting (Zhu, 2012). Despite the attention received by EMD and VMD, its application to the field of

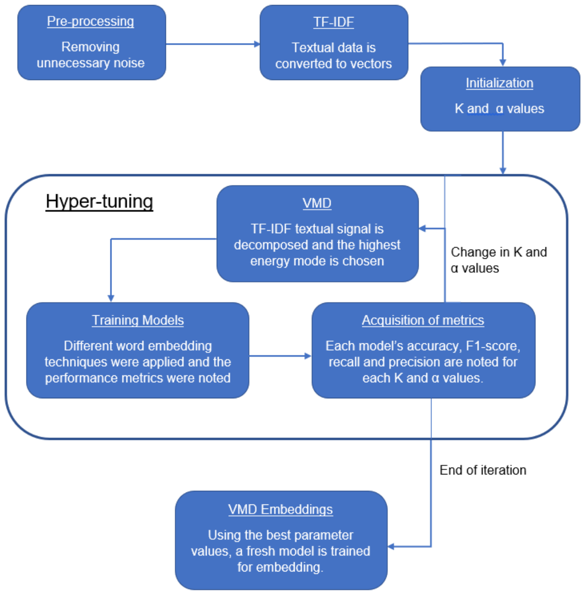


Figure 1: Flowchart depicting the workflow

Natural Language Processing is uncharted.

3 Methodology

The processed data,i.e, after the removal of non-characters, is converted into word vectors or TF-IDF vectors which are collectively summed to form the sentence vectors. These sentence vectors undergo Variational Mode Decomposition. The hyper-parameter tuning for this study is restricted to the number of modes and α-value. The mode with the highest energy is extracted for each sentence and fed into numerous Machine Learning Algorithms to test its performance. Raw Data comprise of numerous punctuation, hashtags, URLs, twitter handles (user-names) and non-characters. The primary step is to remove all of these unnecessary noise to obtain a pure textual data.

3.1 Variational Mode Decomposition

The objective of novel Variational Mode Decomposition (VMD) proposed by Dragomiretskiy Konstantin and Dominique Zosso (Dragomiretskiy and Zosso, 2013),is to obtain modes v_k or sub-signals after decomposing any real valued input signal f, satisfying certain sparsity qualities.(Rilling et al., 2003)

$$\min_{\{v_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad \text{s.t. } \sum_k v_k = f \tag{1}$$

where $\{v_k\} := \{v_1, .., v_k\}$ and $\{w_k\} := \{w_1, .., w_k\}$ symbols representing modes and central frequencies, respectively. The algorithm implementation for Equation (1) purposed by the authors (Dragomiretskiy and Zosso, 2013) :

Algorithm 1: ADMM optimization for VMD

Initialise $\{v_k^1\}\ \{w_k^1\}\ \lambda^1, n \leftarrow 0$;
while

$$\sum_k \left\| v_k^{n+1} - v_k^n \right\|_2^2 / \left\| v_k^n \right\|_2^2 < \epsilon \tag{2}$$

do
n←n+1;
for *k = 1:K* **do**
Update Equation for v_k :

$$v_k^{n+1} \leftarrow \arg\min_{v_k} \mathcal{L}\left(\left\{ v_{i<k}^{n+1} \right\}, \left\{ v_{i\geq k}^{n} \right\}, \left\{ \omega_i^n \right\}, \lambda^n \right) \tag{3}$$

end
for *k = 1:K* **do**
Update Equation for w_k :

$$\omega_k^{n+1} \leftarrow \arg\min_{\omega_k} \mathcal{L}\left(\left\{ v_i^{n+1} \right\}, \left\{ \omega_{i<k}^{n+1} \right\}, \left\{ \omega_{i\geq k}^{n} \right\}, \lambda^n \right) \tag{4}$$

end
Dual Ascent :

$$\lambda^{n+1} \leftarrow \lambda^n + \tau \left(f - \sum_k v_k^{n+1} \right) \tag{5}$$

end

VMD is often used to decompose the signal to attain useful information or denoise the signals via Mode selection. In this study, the textual data obtained from the TF-IDF will be considered as a collection of signals and hence decomposed into Modes. The parameters considered for this study are restricted to the number of modes and α-value.

The number of modes as the name suggests decomposes the given signal into the desired number of modes (K) and α is the balancing parameter of the data-fidelity constraint. The mode with the highest energy was chosen from all recovered modes corresponding to that signal.

3.2 VMD Embedding Formulation

. To evaluate the performance of VMD, modes were extracted from the output of TF-IDF, GloVe and fastText models. Each Tweet or line of Textual data will be converted into numeric vectors via the abovementioned models, and these numeric vectors are endured with VMD to obtain modes. These modes represent the VMD embedding and non-VMD embeddings correspond to the output obtained via the original models.

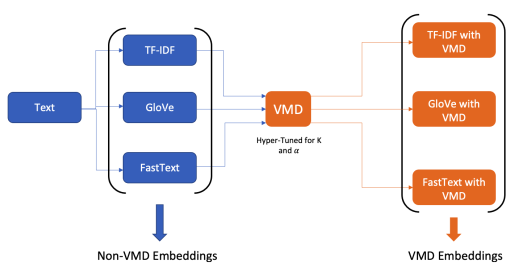


Figure 2: Flowchart depicting the creation of VMD and non-VMD embeddings

3.3 Training Models

. For this study, TF-IDF aided with Tuned-VMD, GloVe and fastText models were used alongside Random-Forest Classifier, Logistic Regression, and KNeighbours Classifier were the algorithms used. The collected dataset was split into a training and testing section. For the train dataset different word embedding techniques were applied and the models were trained to predict using the test data-set and various metrics were loaded without hyper-tuning any of the ML parameters.The performance was noted in terms of accuracy, F1-Score, precision and recall.

4 Results and Discussion

4.1 Hyper-Tuning K-value and α-value

The number of modes were ranged between 2 - 10 and α-value between 2 and 20. Performance of the VMD parameters is directly linked to the performance metrics of the ML model, hence the performance was noted in terms of accuracy, F1-Score, precision, and recall. The table shows that this method shows a high accuracy and also has the best F1 scores.

Fig (3(a)) is the Accuracy, (3(b)) shows the Precision, (3(c)) shows the F1-Score and (3(d)) shows the Recall values; all for varying k-values. Fig (4(a)) is the Accuracy, (4(b)) shows the Precision, (4(c)) shows the F1-Score and (4(d)) shows the Recall values; all for varying α-values.The VMD performed best when K-value and α-value was set to 2.

4.2 Performance of TF-IDF vs VMD aided TF-IDF

. The ML models were trained using both VMD assisted TF-IDF and raw TF-IDF and the parameters of each model were set to default. The metrics obtained from both test results were observed. Fig. (5(a)) compares the metrics of VMD-TFIDF and TFIDF for Logistic Regression, (5(b)) compares the metrics for K-Nearest Neighbours and (5(c)) compares the metrics for Random Forest Classifier. From the experiments performed, Fig. 4 ML models perform better, considering most metrics when trained and tested via VMD assisted TF-IDF.

4.3 t-SNE Comparison

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional data-sets(ref,).

A random list of words were selected from the data-set and the corresponding word vectors obtained via fastText and VMD-TFIDF were dimension reduced via t-Distributed Stochastic Neighbor Embedding (t-SNE) to compare the semantic relationship between words.

Fig. (6(a)), (6(c)) and (8(a)) shows how fastText processes and relates the words, while (6(b)), (6(d)) and (8(b)) shows how VMD-TFIDF processes and relates the words.

Model	Method	Accuracy	F1-score	Precision	Recall	Runtime (s)
Logistic Regression	VMD	0.7175	0.7565	0.6950	0.8314	1.72
	non-VMD	0.5280	0.6910	0.5281	1.0000	1.03
Random Forest	VMD	0.6596	0.7203	0.6359	0.8304	14.3
	non-VMD	0.6095	0.7247	0.5772	0.9734	3.42
KNN	VMD	06595	0.6679	0.6884	0.6486	2.3
	non-VMD	0.4936	0.5164	0.5206	0.5123	2.48

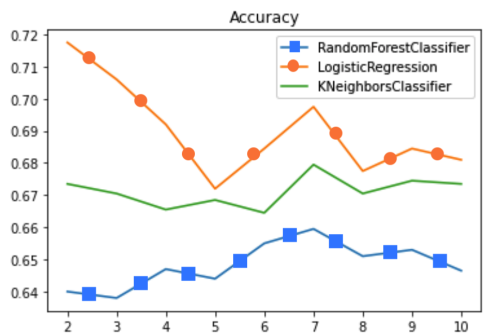
Table 1: Model Comparison TF-IDF

Model	Method	Accuracy	F1-score	Precision	Recall	Runtime (s)
LogisticRegression	VMD	0.7145	0.7369	0.7575	0.7175	0.2373
	non-VMD	0.724	0.7408	0.7472	0.7346	0.339
RandomForest	VMD	0.647	0.7109	0.8219	0.6262	2.24
	non-VMD	0.6735	0.7255	0.8172	0.6523	3.99
KNN	VMD	0.642	0.6313	0.5805	0.6912	2.99
	non-VMD	0.6545	0.6294	0.5558	0.7256	3.53

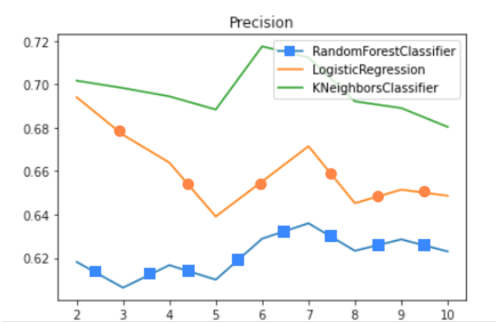
Table 2: Model Comparison GloVe

Model	Method	Accuracy	F1-score	Precision	Recall	Runtime (s)
Logistic Regression	VMD	0.7040	0.7269	0.7462	0.7086	0.587
	non-VMD	0.7160	0.7408	0.7689	0.7149	0.382
Random Forest	VMD	0.6160	0.6821	0.7803	0.6058	2.05
	non-VMD	0.6540	0.7097	0.8011	0.6370	1.76
KNN	VMD	0.6170	0.6348	0.6306	0.6319	4.52
	non-VMD	0.6365	0.6444	0.6240	0.6663	7.52

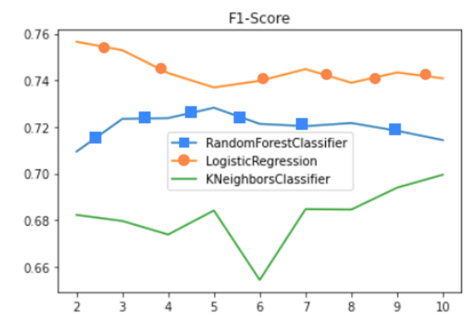
Table 3: Model Comparison fastText



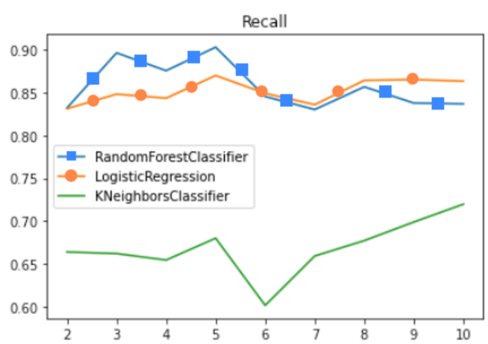
(a) Accuracy



(b) Precision

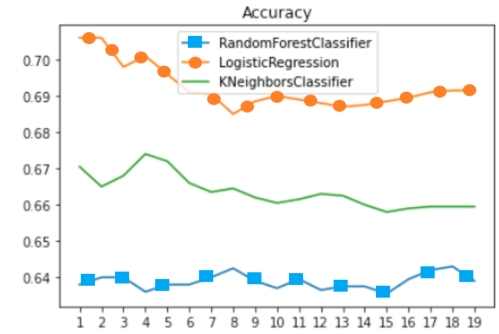


(c) F1-Score

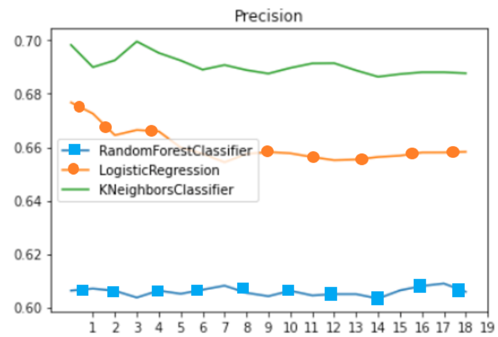


(d) Recall

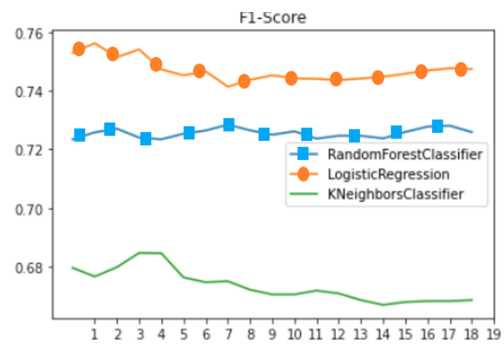
Figure 3: Metric Analysis for various K-values



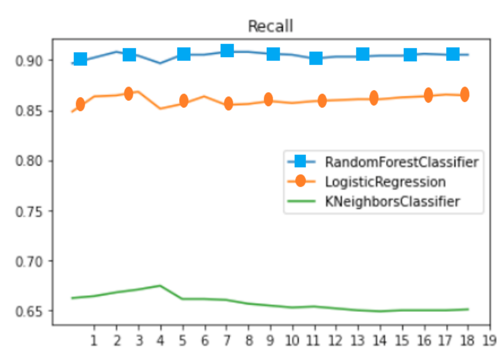
(a) Accuracy



(b) Precision

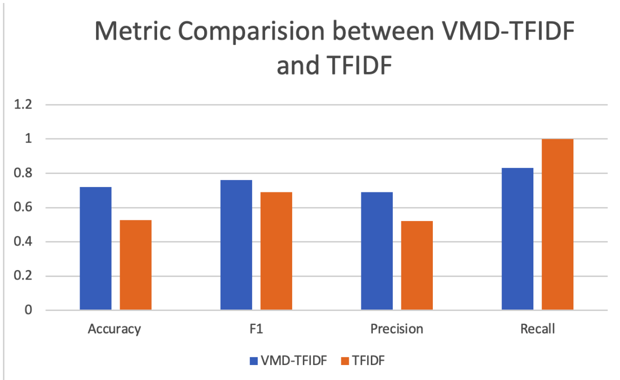


(c) F1-Score

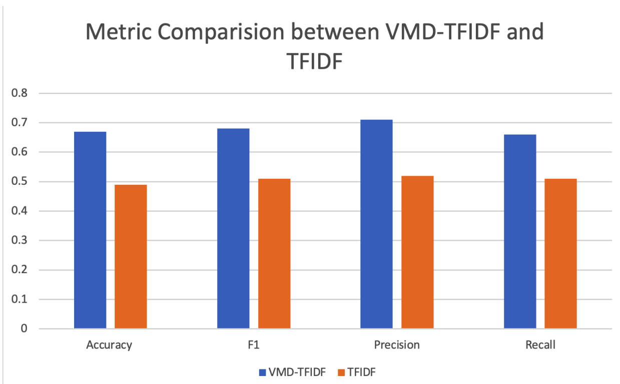


(d) Recall

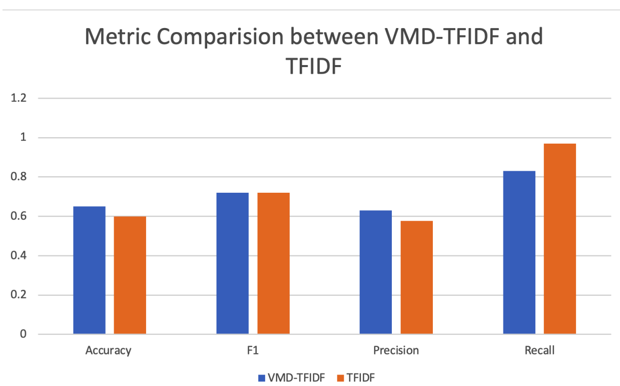
Figure 4: Metric Analysis for various α-values



(a) Logistic Regression



(b) K-Nearest Neighbours



(c) Random Forest Classifier

Figure 5: Metric Analysis for various ML modals with raw TF-IDF and VMD-TF-IDF

From the above results, self-explanatory that the words such as "shelter" and "self-quarantine"; "Today coronavirus" and "corona"; "today covid","people dying" and "hospital"; are closer indicating the better Semitic relation captured by VMD-Embedding compared to fastText.

5 Performance on IMDb Reviews Data-set

The proposed method was tested with the IMDb data-set (Maas et al., 2011), the classification task involves categorising reviews based on the labelled sentiments,i.e, positive,negative,etc. Each reviews were converted into TF-IDF vectors and decomposed with VMD. The mode with the highest energy was extracted and fed into the Classification Model.

5.1 Hyper-Tuning K-Values and α-values

. The number of modes and α-values were ranged between 2 and 6, the corresponding Accuracy metric is visualised (Fig. 8). For each value of K, the value of α was ranged between 2 and 6 resulting in 16 possible combinations. The accuracy metrics from the above-mentioned tuning were plotted. The model performed best with $\alpha = 6$ K = 2.

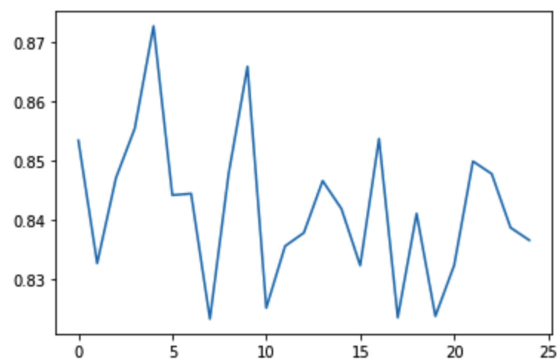
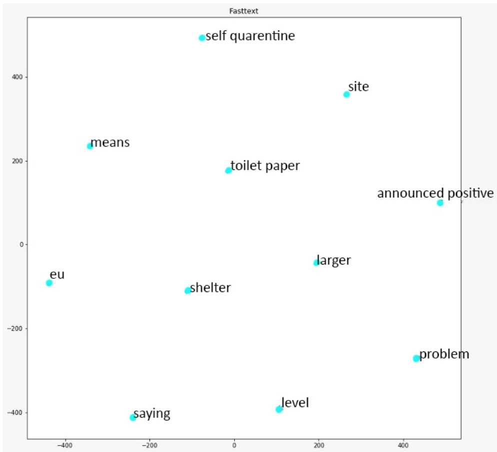


Figure 7: Accuracy Metric Variation with K and α

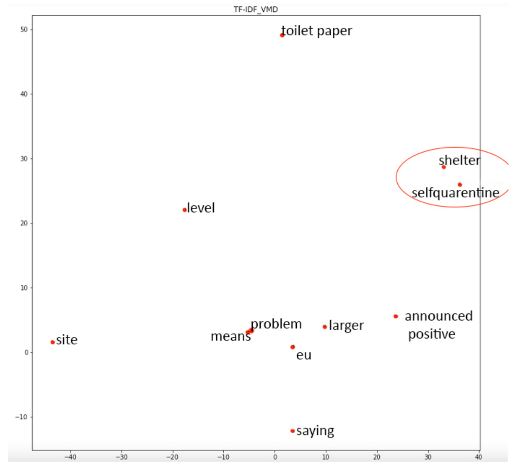
5.2 t-SNE Comparison

. A random list of words were selected from the data-set and the corresponding word vectors obtained via fastText and VMD-TFIDF were dimension reduced via t-Distributed Stochastic Neighbor Embedding (t-SNE) to compare the semantic relationship between words.

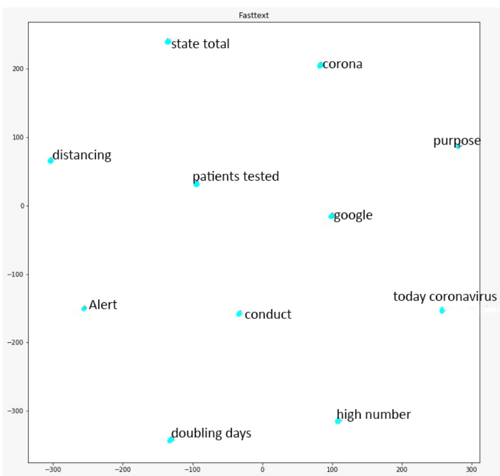
From Fig. 8b, we can observe that words pertaining to emotions and description of the movies, such as *movie good* ,*movie mean* ,*left feeling* ,*hose* ,*engrossed* and *care movie* were closer; Fig. 8d portrays proximity within the words *exudes* ,*hate* ,*inexperienced* and *darkened* in case of VMD-TFIDF convey better context and absorption semantic relation in comparison to fastText.



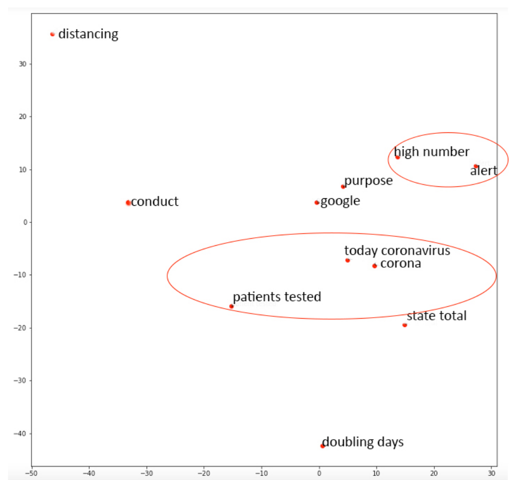
(a) fastText



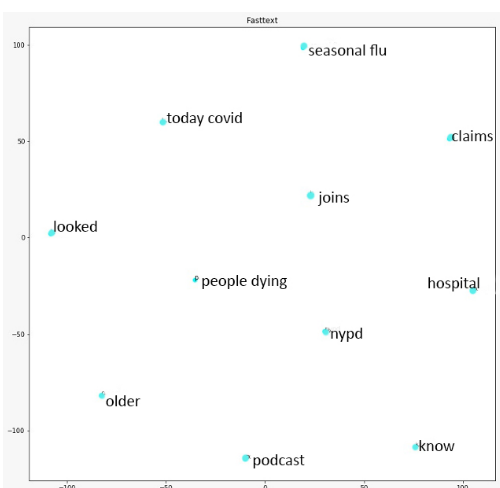
(b) TF-IDF



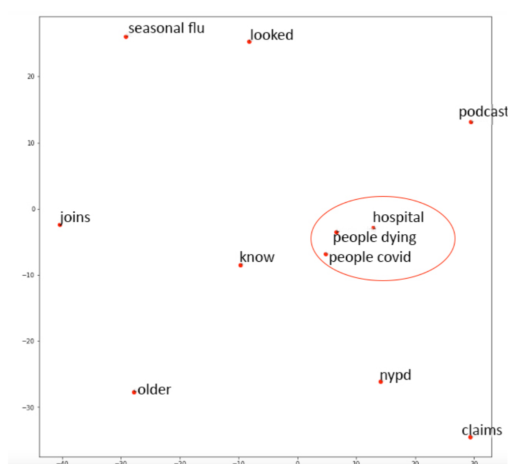
(c) fastText



(d) TF-IDF



(e) fastText

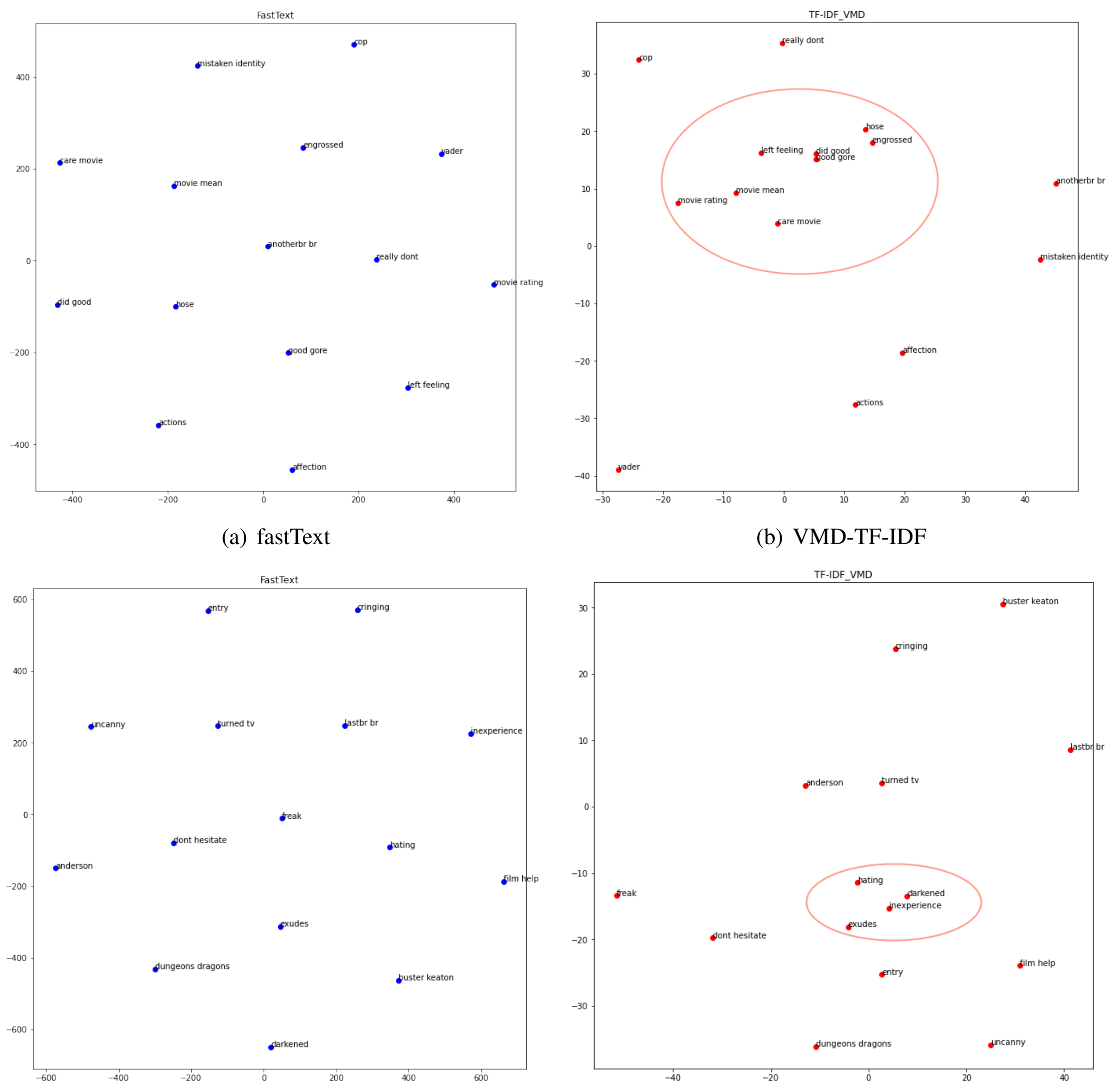


(f) TF-IDF

Figure 6: t-SNE semantic relation of words for VMD-TF-IDF and fastText

Model	Method	Accuracy	F1-score	Precision	Recall
LogisticRegression	TF-IDF-VMD	0.87	0.86	0.85	0.86
	fastText	0.82	0.83	0.81	0.82

Table 4: Model Comparison of fastText and TF-IDF-VMD



(a) fastText

(b) VMD-TF-IDF

(c) fastText

(d) VMD-TF-IDF

Figure 8: t-SNE semantic relation of words for TF-IDF-VMD and fastText

6 Conclusion

The upper hand with the proposed method lies in the transparency of the pipeline. Often Neural-Networks blindly learn the given data-set and are a ”black box” restricting users to tweak parameters to yield a better result. In the proposed method, each step is transparent and accessible by the user. In conclusion, VMD-TFIDF has proven to be at par with Deep Learning Based-embedding due to its ability to capture the semantic relationship among words in a dataset using a simplistic approach.

7 Future Scope

The scope of this study was restricted to testing and analyzing the performance of VMD aided TF-IDF on numerous Machine Learning models and analysing semantic relationships, further experimentation with various Deep Learning models and Neural Networks aided with the VMD could yield results at par or better than the existing Neural Network based embeddings. Different Variants of VMD (Rehman and Aftab, 2019) (Lian et al., 2018) aided with TF-IDF or Deep Learning based embeddings can also be experimented to obtain results proportionate to the existing embeddings, which is beyond the scope of the current study.

References

Barathi Ganesh Hullathy Balakrishnan, Anand Kumar Madasamy Vinayakumar, and Soman Kotti Padannayil. Nlp cen amrita@ smm4h: Health care text classification through class embeddings.

HB Barathi Ganesh, M Anand Kumar, and KP Soman. 2017. Vector space model as cognitive space for text classification. *Notebook for PAN at CLEF*.

Abdel-Ouahab Boudraa, Jean-Christophe Cexus, et al. 2006. Denoising via empirical mode decomposition. *Proc. IEEE ISCCSP*, 4(2006).

Konstantin Dragomiretskiy and Dominique Zosso. 2013. Variational mode decomposition. *IEEE transactions on signal processing*, 62(3):531–544.

Alessio Ferrari, Beatrice Donati, and Stefania Gnesi. 2017. Detecting domain-specific ambiguities: an nlp approach based on wikipedia crawling and word embeddings. In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pages 393–399. IEEE.

C Kalimuthu, EA Gopalakrishnan, and KP Soman. 2019. Compressed air demand forecasting in manufacturing plants using deep learning and variational mode decomposition. In *International Conference on Intelligent Computing and Communication Technologies*, pages 216–225. Springer.

Kais Khaldi, Monia Turki-Hadj Alouane, and Abdel-Ouahab Boudraa. 2008. A new emd denoising approach dedicated to voiced speech signals. In *2008 2nd International Conference on Signals, Circuits and Systems*, pages 1–5. IEEE.

Jijian Lian, Zhuo Liu, Haijun Wang, and Xiaofeng Dong. 2018. Adaptive variational mode decomposition method for signal processing based on mode characteristic, July.

Andrew Maas, Raymond Daly, Peter Pham, Dan Huang, Andrew Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. pages 142–150, 01.

Md Khademul Islam Molla, Somlal Das, Md Ekramul Hamid, and Keikichi Hirose. 2013. Empirical mode decomposition for advanced speech signal processing. *Journal of Signal Processing*, 17(6):215–229.

t-distributed stochastic neighbor embedding. `https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html`. Accessed: 2021-05-30.

Naveed ur Rehman and Hania Aftab. 2019. Multivariate variational mode decomposition. *IEEE Transactions on Signal Processing*, 67(23):6039–6052.

Seyed Mahdi Rezaeinia, Ali Ghodsi, and Rouhollah Rahmani. 2017. Improving the accuracy of pre-trained word embeddings for sentiment analysis. *arXiv preprint arXiv:1711.08609*.

Gabriel Rilling, Patrick Flandrin, Paulo Goncalves, et al. 2003. On empirical mode decomposition and its algorithms. In *IEEE-EURASIP workshop on nonlinear signal and image processing*, volume 3, pages 8–11. Citeseer.

K Soman et al. 2016. Amrita-cen@ msir-fire2016: Code-mixed question classification using bows and rnn embeddings. In *FIRE (Working notes)*, pages 122–125.

Karen Sparck Jones. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21.

Jeffrey Ullman. 2011. *Mining of massive datasets*. Cambridge University Press.

Binwei Weng, Manuel Blanco-Velasco, and Kenneth E Barner. 2006. Ecg denoising based on the empirical mode decomposition. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1–4. IEEE.

Lei Zhou, K.T. Tse, and Yutong Li. 2021. Higher order dynamic mode decomposition of wind pressures on square buildings. *Journal of Wind Engineering and Industrial Aerodynamics*, 211:104545, 04.

Bangzhu Zhu. 2012. A novel multiscale ensemble carbon price prediction model integrating empirical mode decomposition, genetic algorithm and artificial neural network. *Energies*, 5(2):355–370.